



OBJECTIVE

We study action detection in video. We systematically assess the role that human detection can play in action detection. We analyze the statistical properties of intentional movement in action regions and propose a method that leverages such intentional movement without requiring explicit human detection or tracking.

We find that explicitly incorporating human detection performs worse than implicitly incorporating information about the intentional movement of the human. Our action detection training is weakly supervised and bounding-box free.

Trajectories and Intentional Movement

The Puzzle: Action requires the intentional movement, or movement to achieve some active purpose, of an actor. Yet, in works on individual action understanding

like action recognition and action detection, the explicit use of human detection and subsequent processing seems unnecessary. We focus on trajectories since these remain top-performing action features.



The Study: We analyze the spatiotemporal relationship between trajectories and the intentional motion region, by way of annotated bounding boxes. Trajectories are classified accordingly to this spatiotemporal relationship:



AbsPos CenPos FstPos LstPos **PosVol**

All trajectory-points lie in the box The center trajectory-point lies in the box The first trajectory-point lies in the box At least one trajectory-point lies in the box Ratio of box-volume to whole video volume

Observations and Implications

1. Trajectories are tracked well in regions of intentional motion / action; confirming the prerequisite for their use. More than 80% of the trajectories that begin on the action remain on the action for its entirety.

2. Trajectories are elicited in regions of intentional motion: more than 67% of all trajectories occupy actions and hence intentional motion (only one action per video in these data). **3.** Although the majority of trajectories are from the action, on this graph. the relative volume of the action is low. Hence, regions with **Action proposals** are selected as independent connected clusters of trajectories. a high density of trajectories imply intentional motion.

Example Results



Action Detection by Implicit Intentional Motion Clustering Wei Chen¹ and Jason J. Corso² ¹CSE - SUNY at Buffalo, Buffalo, NY; ²EECS - University of Michigan, Ann Arbor, MI

Overview of Action Detection Method







Extract improved dense trajectories

Space-Time Trajectory Graph

poral overlap as motivated by our intentional movement analysis.

the 2D spatial point location \mathbf{x} and the temporal frame f in the video.

$$d^{ij} = \begin{cases} \sum_{k} \|\mathbf{x}_{k}^{i} - \mathbf{x}_{k-o^{ij}}^{j}\|_{2} \\ \infty \end{cases}$$

for the pair by

$$o^{ij} = \min_{\hat{o}} \left(n - \sum_{k} \mathbb{1} \left[f_k^i \right] \right]$$

where 1 is the indicator function. Edges are added among trajectory pairs based on their overlap and distance:



grant (IIS-0845282) and the Army Research Office grant (W911NF-15-1-0354).